DOCUMENT RESUME

ED 067 964 FL 003 698

AUTHOR	Jenkins, James J.
TITLE	Effect of Age, Native Language, and Instruction on Speech Sound Discrimination. Final Report.
INSTITUTION	Minnesota Univ., Minneapolis. Center for Research in Human Learning.
SPONS AGENCY	National Center for Educational Research and Development (DHEW/OE), Washington, D.C.
BUREAU NO	BR-9-0397
PUB DATE	Jun 72
GRANT	OEG-6-9-090397-0076-010
NOTE	26p.
EDRS PRICE	MF-$0.65 HC-$3.29
DESCRIPTORS	Age Differences; Artificial Speech; Auditory Discrimination; *Auditory Perception; Bilingualism; Computers; Electronic Equipment; English; Experiments; *Language; Language Development; Language Instruction; *Language Research; *Language Universals; Phonemes; Phonology; Spectrograms; *Speech; Thai

ABSTRACT

The experiments described in this report seek to investigate the characteristics of speech perception using an approach which considers the development of the perception of "voicing," both as it occurs naturally and as it might occur in the laboratory. Investigating voicing discrimination and perception training among adults, infants, and children, the various experiments seek to determine to what extent the human being is "preprogramed" to make voicing distinctions and to what extent that ability is acquired, whether acquired discriminations are fixed for life or are malleable, whether malleability is a function of age, and whether training techniques can be found which will give an English speaker the ability to make the discriminations the Thai speaker makes. (VM)

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

NCERD FL

Final Report

Project No. 9-0397
Grant No. OEG-6-9-090397-0076 (010)

ED 067964

EFFECT OF AGE, NATIVE LANGUAGE, AND INSTRUCTION ON SPEECH SOUND DISCRIMINATION

James J. Jenkins
University of Minnesota
Minneapolis, Minnesota 55455

June 1972

The research reported herein was performed pursuant to a grant with the Office of Education, U. S. Department of Health, Education and Welfare. Contractors undertaking such projects under Government sponsorship are encouraged to express freely their professional judgment in the conduct of the project. Points of view or opinions do not, therefore, necessarily represent official Office of Education positions or policy.

FL 003 698

U.S. DEPARTMENT OF
HEALTH, EDUCATION, AND WELFARE

Office of Education
National Center for Educational Research
and Development

FILMED FROM BEST AVAILABLE COPY

ACKNOWLEDGMENTS

It is a pleasure to acknowledge the indebtedness of our entire program of research in speech perception to the personnel and facilities of Haskins Laboratories, New Haven, Connecticut. Many Haskins' investigators, especially Alvin Liberman, Frank Cooper and Ignatius Mattingly, have directly shaped both our research and thinking about these problems. Other Haskins personnel, particularly Arthur Abramson and Leigh Lisker, have been generous in providing technical advice and in permitting the use of synthetic stimuli they had generated for their own studies.

Several Minnesota workers have been working guests at Haskins Laboratories during the period of this grant and have drawn support from Haskins' special grant (NIH-71-2420, National Institutes of Health). We are grateful to the people, the laboratory and the National Institutes of Health for the splendid opportunities made available to us.

Second, we must acknowledge our debt to the Center for Research in Human Learning at the University of Minnesota, the matrix in which we live our professional lives. The Center, supported by the National Science Foundation (GB-17590), the National Institute of Child Health and Human Development (HD-01136), and the Graduate School of the University, has housed us and our research efforts during the grant period. Many of the people who have been active in this research are predoctoral fellows and associates of the Center and we are doubly indebted for their help and participation.

For special help with technical problems (such as calibration of the synthesizer) and for his participation in the Speech Research Group we are indebted to Dr. Mark Medress, Systems Design Engineer with the Signal Processing Department at UNIVAC. Contact with the UNIVAC group, which is concerned with both speech recognition and synthesis, has been helpful and stimulating to our project.

Of course, we recognize with gratitude the program of basic research sponsored by the Office of Education which supported the whole endeavor. We have further been fortunate in having Dr. Monte Penney as our "overseer" from OE. He has been a stimulating and enthusiastic coordinator and we have enjoyed this contact with him, and through him, with the agency.

Finally it is a pleasure to record our debt to Miss Kathleen Casey who not only took care of our secretarial problems but also managed the business aspects of the grant.

James J. Jenkins
For the Research Personnel

TABLE OF CONTENTS

GENERAL INTRODUCTION 1

THE RESEARCH

I. Introduction 3

II. Developmental Studies 6

III. Training Studies with Adult Subjects . . 9

IV. Bilingual Studies 13

RESEARCH FACILITIES

I. Speech Synthesis Laboratory 16

II. Infant Testing Laboratory 20

APPENDIX I. Dissemination Activities 22

APPENDIX II. Personnel 23

INTRODUCTION

More than three years ago, I submitted a proposal suggesting an investigation of some characteristics of the perception of speech. The proposal pointed out that many interesting problems could be examined by an experimental approach to the development of the perception of "voicing", both as it occurred naturally, and as it might be made to occur in the laboratory. The proposal also pointed out that "voicing" is a universal or nearly universal feature of the phonology of language, that varying degrees of voicing can be synthesized, and that techniques exist for examining the identification and discrimination of synthetic speech.

Commonly, adults accurately discriminate degrees of voicing at only one or two points along the continuum of variation in voicing. These points are those marking the separation of phonemes along that continuum in their own natural language. But, since not all languages make the same separations in the continuum, it is clear that the way people <u>are</u> is not necessarily the way they <u>must</u> be. To what extent is the human being "preprogrammed" to make voicing distinctions, and to what extent is his ability acquired? If a given set of discriminations is acquired, are they fixed for life, or are they malleable? Is malleability a function of age? Can training techniques be found which will give an English speaker the ability to make the discriminations the Thai speaker makes?

These are very ambitious questions and, of course, the two years of research we have conducted have only begun to answer them. It is a tribute to my students and associates, however, that we have attacked at least some aspect of each of these problems. Aided by our friends and collaborators at Haskins Laboratories, we were able to develop synthetic materials for our problems and begin the research while our own synthesizer was being readied. While Steve Mullen struggled with the many trials involved in that endeavor, Pat Yonas, Dennis Doty, Winifred Strange, and J. Richard Barclay began laboratory work with infants, children and college students on the problems sketched in the proposal. Later Ron Howes took up where Mullen left off and became our computer-synthesizer genius. About that time, Robert Verbrugge joined us to undertake the task of developing a synthesis-by-rule program for the laboratory.

Since most of our money went into equipment, technicians, subjects and supplies, it follows that most of the research assistants were unpaid; and so they were. To all of them special thanks must be given, and most especially to Winifred Strange who organized (and provided the spirit for) our speech perception research group and who kept the laboratory on some semblance of schedule. In addition, of course, she did the major part of the training research.

What do we know now that we did not know three years ago? First, we confirmed the finding that infants do, indeed, respond differentially to speech sounds. That is, there is some likelihood that some speech distinctions are already "built-in" for the infant. Second, we found that 10- and 11-year-old children did not respond dramatically to training on distinguishing the voicing continuum; in fact, they scarcely changed at all. Further, we found that college students did not respond well either, though we tried many forms of training experience. Some helpful notions emerged, however, along with the hint that early experience with several languages may be a great advantage in learning phonological distinctions later. Finally, we examined bilinguals and confirmed the findings of Abramson and Lisker with respect to their earlier research on Spanish and Thai speakers.

The report that follows sketches in a little more detail these major findings: First, there is a brief introduction to the notion of voicing and a summary of the findings that led us to undertake the research. Section II then treats with the studies of infants and children. With infants, adaptation techniques were used to see whether or not the babies detected the differences that adult listeners heard in the stimuli. With children, deliberate training techniques were employed. Section III deals with efforts employing several different techniques to train college students to hear the voicing distinction employed in Thai but not in English. Finally, Section IV reports the bilingual studies of voicing identification and discrimination.

Because a good deal of emphasis went into the development of laboratory facilities for further speech perception research, a special section is devoted to the facilities and the resources developed under this grant. The first section on facilities deals with the speech synthesis laboratory and its hardware and software. The second section describes the laboratory for research on infant speech perception.

Appendices at the end of the report record the activities of personnel on the grant in dissemination of information on speech perception, and list the personnel who worked on the project during this period of support.

James J. Jenkins
Principal Investigator

THE RESEARCH

I. Introduction

Through spectrographic analysis of speech sounds from eleven languages, Lisker and Abramson (1964) found a single acoustic continuum that underlies both the prevoiced-voiced distinction in languages such as Thai and Spanish, and the voiced-voiceless distinction which differentiates English voiced /b/, /d/, and /g/ from voiceless /p/, /t/, and /k/. Initial stop consonants can be organized along a continuum where the relation between the onset of the second formant, F2 (corresponding articulatorily to the release of closure), and the onset of the first formant, F1 (glottal pulsing), varies from voicing lead (F1 preceding F2), through simultaneous onset, to voicing lag (F2 preceding F1). This continuum is called Voice Onset Time (VOT). Prototypes tend to fall into one of three groups: (1) 50 to 150 msec. "voicing lead," (2) 0 to 30 msec. "short voicing lag," or (3) 50 to 110 msec. "long voicing lag."

English differentiates the long lag group, voiceless (aspirated) phonemes [ph], [th], and [kh], from the other two groups, which together form the class of voiced phonemes /b/, /d/, and /g/. Spanish, on the other hand, distinguishes phonemes with voicing lead (prevoiced consonants) from those with short voicing lag (unaspirated [p], [t], and [k]). For Thai, all three groups correspond to phonemic categories: prevoiced, voiced, and voiceless stops, respectively.

Using synthetic speech sounds generated by a computer-driven speech synthesizer, Abramson and Lisker (1967) tested speakers of Thai, Spanish and English on their ability to discriminate acoustic differences along the VOT continuum. Results generally showed that discriminability of differences in VOT for sounds drawn from within one phonemic category was very poor; sounds varying by the same amount acoustically, but which are from different phonemic categories, were discriminated quite accurately in a sequential oddity discrimination task. For English speakers, discrimination along the VOT continuum was good only for those values that fall at the boundary between long lag and simultaneous or short lag. Discrimination of the VOT values between voicing lead and short lag where there is no phonemic boundary in English was no better than for any of the other values within phoneme categories. Thai speakers, on the other hand, showed two regions of accurate discrimination corresponding to the two phonemic boundaries in their language. The results for Spanish speakers were somewhat less clear, but showed a tendency toward a "peak" of relatively more accurate discrimination in the prevoiced-voiced phoneme boundary region. Thus, it seems that discrimination of voice onset time is dependent on the acoustic differences being "linguistically significant" in the native language of the subject.

The dependency of the perception of acoustic parameters on the specific language experience of speaker-hearers raises several interesting questions about the development and modification of speech perception processes. Some of these questions were investigated in the research reported here. The first set of studies dealt with the course of development of perceptual boundaries in infants and children of English-speaking background. Pre-linguistic infants were tested to determine if they discriminated VOT in a discontinuous manner. A study with pre-adolescent children was conducted to determine if their discrimination was similar to that of adults. They were then given discrimination training to investigate the acquisition of a new phoneme distinction, such as the Thai prevoiced-voiced contrast.

A second series of studies asked whether adult English-speaking subjects could learn to perceive the VOT continuum as Thai speakers do. The effectiveness of several training techniques in modifying discrimination performance was investigated.

Finally, two studies with bilingual subjects were conducted to further explicate the role of linguistic experience in the perception of Voice Onset Time. Spanish-English and Thai-English bilinguals were tested.

Stimulus materials for this research were generated at Haskins Laboratories on a parallel-resonance synthesizer, using control parameters developed by Abramson and Lisker. A series of Consonant-Vowel syllables was constructed which varied in VOT from 150 msec. voicing lead, through simultaneity, to 150 msec. voicing lag in 10 msec. steps. In referring to the stimuli, prevoiced variants are designated as negative; postvoiced variants have positive values. Two such series were constructed: the labial series (b-p) and the dental series (d-t). They are identical with respect to VOT and differ only in the form of the F2 transition which determines the place of articulation.

The basic test of discrimination used in studies with adults and children is called the Oddity Test. Pairs of VOT variants are drawn from the series such that each differs by the same amount, (e.g., 30 msec.: -150/-120 VOT, -140/-110 VOT, etc.). For each comparison pair, sequential oddity triads are constructed by reduplicating one of the variants and arranging the three in all possible permutations. The subject's task is to detect which of the three stimuli is the different or "odd" one. Discrimination functions plot the percentage of correct judgments for equal interval comparison pairs, and thus represent the relative discriminability of VOT across the continuum.

The studies reported below utilized a modification of this basic discrimination measure which incorporates weighting factors from subjects' confidence ratings of their discrimination judgments. This technique was found to increase the reliability of the

discrimination measure computed over the same number of discrimination judgments, (Strange and Halwes, 1971).

Discrimination functions obtained in this fashion can be compared with results of identification tests, in which the subject's task is to assign phoneme labels (letters) to VOT variants presented one at a time in random order. If "peaks" of relatively more accurate discrimination occur in the boundaries between labeled phoneme categories, perception is said to be "categorical."

II. Developmental Studies

A. Discrimination of Voice Onset Time by 2-, 6-, and 10-Month Olds

This research explores the developmental implications of the findings regarding the perception and production of initial stop consonants by adults.

These findings suggest the joint operation of biological and experiential factors: Production of VOT differences may be biologically constrained, but the perception of such differences clearly depends on linguistic experience. However, discriminability at certain points on the VOT dimension may be influenced by innate factors as well.

A study by Patricia Yonas asks two main questions: Are infants sensitive to VOT differences, and, if so, when do they begin perceiving them categorically as linguistically relevant? A within-subjects repeated measures design was used to compare the discriminability of four VOT comparison pairs. The members of each pair are separated by a 40 msec. difference in VOT. One pair crosses the boundary between the English phonemes /b/ and /p/ (0 VOT/+40 VOT). Two different pairs represent phonemic distinctions in languages other than English (-40 VOT/ 0 VOT, -80 VOT/-40 VOT), and the fourth pair represents no known phonemic contrast optimally (+40 VOT/+80 VOT). (An adult speaker of English would discriminate the first contrast only, hearing the other pairs as indistinguishable variations of /ba/ or /pa/.)

Heart-rate change was used as the index of discrimination. With repeated presentation of a stimulus, the initial heart-rate deceleration habituates, to recur or dishabituate with subsequent stimulus change. Each of 72 infants (of English speaking parents) either 2-, 6-, or 10-months old, heard each of the four stimulus pairs (one a day in counterbalanced order), receiving eight trials of stimulus A followed by two trials of stimulus B. (A trial consisted of ten repetitions of the stimulus separated by a 1-second interstimulus interval; the intertrial interval was 15 seconds.) Differences in both magnitude and latency of the decelerative response immediately before and after stimulus change indicate discrimination. A less stringent test is provided by comparing responses of experimental subjects during the change trials with those of a control group (N = 18) which received ten trials of stimulus A.

Three general outcomes are possible: (1) Infants may discriminate all four contrasts equally well; that is, they discriminate VOT differences but may not be processing the sounds as speech. (2) Infants may discriminate the three language-relevant contrasts better than the "nonphonemic" contrast, suggesting that biological constraints on perception influence the placement of

phoneme boundaries and that perceptual learning is a matter of acquiring equivalences among previously distinguishable sounds. (3) Infants may discriminate the English contrast and no other (or better than any other), indicating that linguistic experience has had an effect. The age at which this occurs is of major interest (e.g., is it before or after babbling begins?), as is the pattern of discrimination preceding this adult-like perception.

Initial results showed little reliable discrimination for the 2-month-old infants for any of the comparison pairs. The 6- and 10-month-old subjects showed good discrimination for the English contrast (0 VOT/+40 VOT). There was some evidence for discrimination by some subjects of the non-English phonemic comparison pairs. Further testing and analysis is currently underway to obtain clearer answers to the questions of the effects of linguistic experience on perception.

B. Child Training Study

Dennis Doty turned to the question of whether preadolescent children could learn new identifications and discriminations of labial consonants varying in Voice Onset Time. Ten- and 11-year-olds were chosen for this training study because they were old enough to undergo a long and demanding series of training sessions, yet young enough to fall within Lenneberg's critical period for language acquisition. Identification and discrimination pretests showed that these children perceived the VOT continuum exactly as English speaking adults do, i.e., categorical with respect to a single /b/-/p/ boundary at about +25 msec. VOT.

The oddity discrimination task was used in the training procedure. Three stimuli were played to the subjects in rapid succession, two of which were the same, and one of which differed from the other two by 30 msec. VOT. After subjects had indicated which stimulus they thought was the different one, they were told the correct answer, rewarded with tokens exchangeable for money if they were right, and presented with the triad once again so that they could try to hear the correct distinction. Every effort was made to facilitate learning: The subjects were told about the nature of the VOT continuum, they were given exaggerated examples of prevoiced and postvoiced variants, they were encouraged to verbalize what they thought were the crucial cues, and they were given graphs of their past performance and told after each test trial from where on the graph of the VOT continuum the stimuli had been drawn. Training stimuli were drawn from throughout the continuum; and it was hoped that the discrimination training would result in the creation of new identification categories. Each discrimination pair was presented 36 times during training in 18 sessions extending over three months. Despite the long, tedious nature of the training procedure, the system of rewards for correct responses maintained the subjects at a high level of motivation, and they were actually disappointed to see the study end.

After training, the subjects were given a discrimination posttest and two identification posttests: one in which they were to use the two categories they had used in pretesting, and one in which they were to use four individually chosen categories. All subjects continued to identify a sharp /b/-/p/ boundary, but had only moderate success in creating other boundaries. There was no agreement between subjects on where these new boundaries should be placed, and no tendency to place a boundary where the Thai prevoiced-voiced boundary occurs. There was little correspondence between the placement of new boundaries and post-training discrimination performance. In other words, perception did not appear to remain categorical. All subjects showed some minor improvements in discrimination at various spots along the continuum, but the only place where all subjects showed marked improvement was just on the prevoiced side of the original English /b/-/p/ boundary. The effect of this improvement was to create for all subjects an area of high discrimination roughly 50 msec. wide and centered at 0 msec. VOT. The synthetic stimuli used in this study, like real speech stimuli, have two correlated cues which are non-continuous at this point on the continuum. These discontinuities are (1) presence versus absence of F1 in isolation (a "voicing bar") and (2) presence versus absence of non-periodic glottal excitation (F2 is non-periodic preceding F1 onset in postvoiced variants). It was concluded that these physiologically based discontinuities in the voicing continuum are the basis of a very strong phonetic universal corresponding to the voiced-voiceless distinction in both Thai and English. The study provided no evidence that there was a prepotency to further differentiate the continuum at the Thai prevoiced-voiced boundary for English speaking children.

III. Training Studies with Adult Subjects

A series of studies by Winifred Strange tested the effects of three kinds of training techniques on the discrimination of the prevoiced-voiced distinction by adult English-speaking subjects. The basic design of the experiment included pretests to establish pretraining discrimination performance, repeated training sessions, and discrimination posttests. The measure of the effectiveness of training was thus the improvement by each subject over his pretest performance.

A. Oddity Discrimination Training

The first training technique involved repeated practice in the oddity discrimination task with feedback after each discrimination judgment. The purpose of this study was to ascertain whether subjects would show differential improvement in their ability to discriminate in the region of the Thai prevoiced-voiced boundary when given equal amounts of practice along the full labial VOT continuum. Four undergraduate students were pretested on identification and discrimination of the labial VOT series. All showed characteristic peaks of accurate discrimination in the English /b/-/p/ boundary.

Training procedures were similar to those used by Doty described above. After the subject listened to a discrimination triad and made his response, including a confidence rating, the experimenter responded "right" or "wrong". (The correct response was not indicated if the subject erred.) Subjects continued through the task at their own rate. Subjects completed a total of 20 repetitions of the training task in five sessions taking place over a three-week period. Post-training discrimination tests, in which no feedback was given, were administered immediately after the completion of training and again after approximately four months.

Results of the comparison of pretests and posttests given immediately after training showed no significant improvement in discrimination in the Thai prevoiced-voiced boundary region. However, all subjects' posttest functions indicated substantial improvement in the discrimination of pairs to the immediate prevoiced side of the original English peak, similar to that found by Doty with 10- and 11-year-old subjects. This change over pretest performance was maintained somewhat over a four-month period. The delayed posttests also showed improvement over pretests for these pairs, albeit less than on immediate posttests.

B. Identification Training

The second training study was designed to explicitly test the idea that discrimination of speech sounds is a function of the categories which subjects identify as distinct "linguistic"

classes. In order to focus the subjects' attention on the distinction of interest (i.e., the Thai prevoiced-voiced contrast) only the stimuli which were identified as within the voiced category on pretests were used in this study (-100 VOT through +10 VOT, inclusive). Three subjects were administered discrimination tests of the labial and dental partial series for comparison with post-training performance.

Training consisted of two tasks. First, subjects learned to differentially label the two endpoints of the dental series (i.e., -100 VOT and +10 VOT) to a strict criterion. Then, they were presented the intermediate stimuli of the series one at a time in random order, and assigned one of the labels to each stimulus. In other words, they were required to categorize the series into two distinct classes, using as references the labeled endpoints. No feedback was given during this latter task; subjects were shown their identification functions after each session. They continued the identification task until performance stabilized. Subjects completed a total of 90 identification trials in eight sessions over a period of six weeks.

Immediately following the final training session, discrimination posttests of both the dental (trained-on) and labial (transfer) series were given. In addition, subjects completed an identification test of the labial series.

The three subjects varied with respect to how well they were able to differentiate two classes by the end of training. Two of the three subjects showed identification boundaries which were consistent and approached those of Thai speakers in sharpness. The third subject produced a very gradual slope in the crossover from one class to the other.

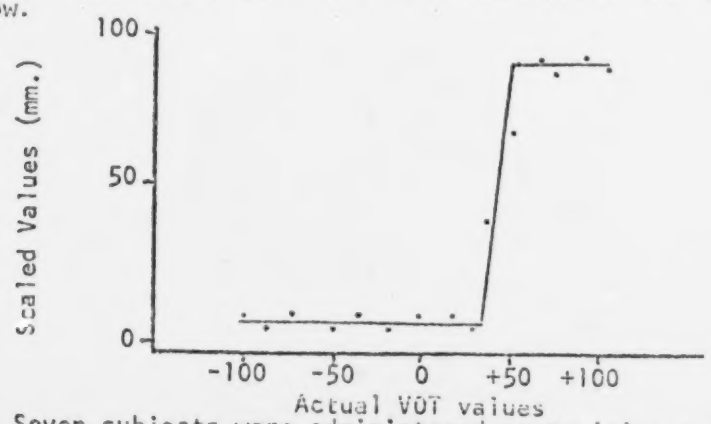
Results of the pretest-posttest discrimination comparison reflected the same variability. For the two subjects whose boundaries were well defined, posttest discrimination functions showed peaks of relatively more accurate discrimination for pairs which were drawn from different established categories. Discrimination of pairs taken from within a category showed no improvement. The third subject, while indicating a slight overall improvement in discrimination over pretest performance, produced no peak of accurate discrimination in the boundary region. It was concluded that discrimination of these speech sounds did show a definite relationship to identification performance, i.e., perception tended to be "categorical" for those subjects who had succeeded in establishing well-defined categories during the training.

Discrimination and identification tests of the labial series were included to test for possible transfer effects of the training on discrimination of the same feature distinction in a different phoneme environment. (In other words, did subjects learn a new feature distinction, or was learning specific to the phonemes used

in training?) No transfer effects were found in the post-training discrimination results. None of the subjects showed improvement over pretest performance. Identification tests of the labial series also indicated no transfer of training; subjects were not able to differentiate the series into distinct classes. It appeared that the change in perception of the VOT continuum was specific to the stimuli presented during training.

C. Training in Continuous Scaling

For the third study, a continuous scaling test, developed by Port and Yeni-Komshian (1970) was utilized. In this task, subjects are presented stimuli of the VOT series one at a time and respond by indicating on a continuous linear scale their similarity to reference endpoints (-100 VOT and +100 VOT). Port and Yeni-Komshian found that English-speaking subjects produce categorical functions in this task analogous to their performance on discrimination tests. That is, all VOT variants within the same labeled phoneme category are judged as equidistant from the endpoint references, producing a function such as the one shown below.



Seven subjects were administered pretraining tests in scaling the labial and dental VOT continua. They then completed oddity discrimination pretests of the partial labial and dental series (-100 VOT to +10 VOT), identical to those used in the previous experiment.

Training consisted of first, learning to differentially label the endpoint reference of the partial series (i.e., -100 VOT and +10 VOT). After reaching a strict criterion, subjects then scaled intermediate VOT variants of the series with respect to these reference stimuli. Before each training session, they heard the series in order from -100 VOT through +10 VOT twice in succession. Twenty-one scaling tests were completed in seven sessions over the period of three weeks. Three subjects trained on the labial series, four on the dental series.

Post-training discrimination tests of the two partial series were administered for comparison with pretest performance. Finally, subjects scaled the full VOT continua, both labial and

dental, with reference to the original endpoint references.

Over the course of training sessions, subjects varied greatly in their scaling performance. Some subjects showed gradual and consistent improvement in their ability to continuously scale the intermediate VOT variants, as indicated by increased positive correlations of judged and actual "distance" from endpoint references. Other subjects produced inconsistent results over training sessions, and two of the seven subjects showed a deterioration of scaling performance over sessions. Comparisons of pretraining and post-training discrimination functions were equally heterogeneous. While some subjects showed overall improvement, there was no relationship between scaling performance and discrimination results. It appeared that, while some subjects were capable of more continuous perception of the VOT series as a function of practice, their performance was quite unstable, and the training had no clear effect on discrimination of VOT as tested by the conventional oddity procedure.

D. General Conclusions from Training Studies

The series of training studies produced somewhat unsatisfying results with respect to the original question of whether adult English speakers could learn the Thai prevoiced-voiced distinction. The training method which met with the most success was the identification procedure used in the second study. However, even in this case, all subjects did not succeed in the task. Also, there was no indication that what subjects did learn was transferable to any other speech sounds than those specifically trained on.

Some implications, both for the basic understanding of speech perception processes and for practical concerns of foreign language teaching might be drawn from the limited success of these experiments. First, the failure to easily modify subjects' perception by any of the techniques utilized argues against a simplistic, operant conditioning explanation of the phenomenon of categorical perception such as that proposed by Lane (1965). The relatively great success of the identification training supports the general theoretical notions of the Haskins researchers that categorical perception is a function of special processes employed in the rapid identification of highly encoded speech sounds.

This suggests that any modification of the perception of these kinds of speech stimuli must, in some way, involve these processes. That is, subjects must learn to "cut up the acoustic world of speech" into new, qualitatively distinct categories of perceptual "objects."

IV. Bilingual Studies

A. Spanish-English Bilingual Study

Abramson and Lisker have shown that monolingual speakers of Puerto Rican Spanish place the boundary between prevoiced /b/ and unaspirated /p/ at a less post-voiced location than the English b-p boundary. In identifying synthetic labials, the Spanish speakers placed their boundary at about +10 VOT, and the English speakers placed their boundary at about +25 msec. The purpose of this study by Doty was to see if Spanish-English bilinguals would be able to identify and discriminate both Spanish and English phoneme distinctions in a series of synthetic labials.

The subjects were four native Columbians, who spoke English fluently as their second language, and one native American graduate student in Spanish. The subjects were given seven presentations of each of the 31 stimuli spaced 10 msec. apart from -150 to +150 msec. VOT in random order and told to identify them into as many categories as they could. The American subject and the one Columbian who had had training in phonetics divided the continuum as a phonetician would into three categories: [b], [p], and [ph]. However, the boundary between [p] and [ph] was placed at +50 to +60 msec. VOT, instead of at about +25 msec. where the Abramson and Lisker data suggest it should be placed. It would appear, therefore, that phonetics courses teach an inaccurate, overly aspirated version of the aspirated [ph]. The remaining three Columbians divided the continuum into just /b/ and /p/, placing the boundary closer to the English location of +25 VOT than to the Spanish location of +10 VOT.

All subjects were given pairs of the stimuli they had heard in the identification task for discrimination in an oddity discrimination task. The stimuli to be discriminated differed by 20, 30, or 40 msec. All subjects showed categorical discrimination with respect to the categories they had identified; that is, they showed nearly perfect discrimination at phoneme category boundaries, but almost no discrimination within boundaries. The two subjects who had identified a [p]-[ph] distinction showed improved discrimination at the location, albeit inaccurate, where they had identified this boundary, although discrimination at their [p]-[ph] boundary was not as good as at their [b]-[p] boundary. Two of the Columbians produced discrimination peaks which were broader than would be predicted on the basis of their identification performance. These two subjects showed improved discrimination from 0 to 50 VOT.

This study thus provided some evidence that new identifications and discriminations could be learned along the VOT continuum. The discrimination data were hard to interpret, however. Abramson and Lisker were unable to obtain clear discrimination data from their Puerto Rican monolinguals. Unfortunately, Spanish monolinguals

could not be found in the Minneapolis area, so that good monolingual comparison data remained unavailable.

B. Thai-English Bilinguals

In conjunction with the training studies reported above, three Thai-English bilingual students were tested by Strange as a partial replication and extension of Abramson and Lisker's (1967) research. These subjects completed identification and discrimination tests on the labial VOT continuum, identical to those used for pretesting in the training studies. In addition, they were administered identification and discrimination tests of the partial (-100 through +10 VOT) labial and dental continua. On discrimination tests, subjects were required to rate each of their responses by the three category rating technique described above.

Results of the initial labial discrimination tests yielded results similar to those found by Abramson and Lisker. That is, functions showed two peaks of relatively more accurate discrimination for pairs drawn from different phoneme classes. However, discrimination of pairs in the prevoiced-voiced boundary was inferior to that of pairs in the voiced-voiceless boundary, both in terms of correct responses and in confidence of judgments. This corroborates the trend shown in Abramson and Lisker's data; that the second boundary for Thai speakers is less sharply defined. When tested on the partial continua, which contained only prevoiced and voiced VOT variants, identification boundaries were sharp and discrimination functions revealed clearer peaks of relatively accurate discrimination. It was concluded that, while the data indicated that Thai speakers did indeed perceive the VOT series as consisting of three distinct phoneme classes, in contrast to English speakers, the prevoiced-voiced distinction was not as salient a contrast as the shared English-Thai voiced-voiceless distinction. This suggested that the acoustic cue of Voice Onset Time might not be the best or only parameter distinguishing this phoneme feature in natural speech.

REFERENCES

Abramson, A.S., & Lisker, L. Discriminability along the voicing continuum: Cross-language tests. Proceedings of the 6th International Congress of Phonetic Sciences, Prague: 6-7 September 1967.

Lane, H. The motor theory of speech perception: A critical review. Psychological Review, 1965, 72-4, 275-309.

Lisker, L., & Abramson, A.S. A cross-language study of voicing in initial stops: Acoustical measurements. Word, 1964, 20, 384-422.

Port, D., & Yeni-Komshian, G. H. Use of a scaling technique in the perception of stop consonants along a voicing continuum. Unpublished manuscript, 1970.

Strange, W., & Halwes, T. Confidence ratings in speech perception research: Evaluation of an efficient technique for discrimination testing. Perception and Psychophysics, 1971, 9, 182-186.

RESEARCH FACILITIES

1. Speech Synthesis Laboratory

The Speech Synthesis Laboratory is used for the production of synthetic speech stimuli for experimental studies and for research in techniques of speech synthesis using formant synthesis and synthesis-by-rule. The laboratory contains a Digital Equipment Corporation PDP-8/L computer interfaced with a Glace-Holmes speech synthesizer and a variety of peripheral equipment. The peripheral equipment includes: (1) a model ASR-33 teletype, (2) a Digital Equipment Corporation (DEC) type PR8/L high speed paper tape reader, (3) a DEC type DF32 disc storage unit, (4) an audio patch panel, (5) an Ampex model AG500 tape deck, and (6) an amplifier/speaker unit. The program software available for use with this equipment allows the user to control efficiently the types of speech sounds which may be produced, and includes a formant synthesis program and a program for synthesis-by-rule.

Hardware:

The PDP-8/L computer is a twelve-bit word machine with 8192 words of core memory and is capable of performing 312,500 arithmetic operations per second (1). It is used for running all programs necessary for operating the computer system, for controlling the speech synthesizer through the synthesizer interface, and for storing the control parameters which are loaded into the synthesizer interface to produce speech sounds.

The speech synthesizer interface includes eleven six-bit digital-to-analog (D/A) converters, which translate the digital control parameters stored in the computer memory into eleven analog signals which control the synthesizer. A one kilohertz crystal clock provides a highly stable time base which is used to determine the rate at which control parameters are loaded into the interface, and, therefore the rate at which speech sounds are produced by the synthesizer. A cue interrupt circuit is provided which allows external control of the initiation of synthesis.

The Glace-Holmes terminal analog speech synthesizer contains several filter circuits which act in parallel on the output of either a noise source or a broad spectrum pulse source. The frequency and amplitude of these circuits, or resonances, as well as the selection of the source is dynamically controlled by the eleven analog input signals. The output of these resonances is then combined and amplified, producing a net sound with formant characteristics similar to natural speech (Glace, 1968).

Speech sounds are synthesized by specifying control parameter values in sequential time units or frames of 1 to 200 msec. duration. The parameter values for the time frames are stored in a buffer in the computer. During the generation of the speech sound, the parameters are loaded into the interface frame by frame and

transformed into analog signals which drive the synthesizer.

The primary input/output device for the computer is the ASR-33 teletype. The user controls the computer by typing data and commands on the teletype keyboard and receives information from the computer through the teletype printer. A paper tape reader and punch, which are mounted on the teletype, can be used for input and output of information on punched paper tape.

The DEC type PR8/L high speed paper tape reader is capable of reading tape at up to 300 frames per second and is used for high speed input of programs and data which are stored on this medium.

The DEC type DF32 disc storage unit which can store up to 32,768 twelve-bit words provides high speed random access storage of programs and data (1). The software associated with this unit allows the user to store programs and data, and later recall this information through commands typed on the teletype keyboard. This allows the user to control the flow of information through the computer without leaving the teletype keyboard--a method which is superior to handling paper tapes or other storage mediums.

The audio patch panel and the Ampex model AG500 tape deck allow the user to record the speech sounds produced by the synthesizer for use in experimental studies or for later playback in an environment where the speech sounds may be accurately evaluated. The recordings may be of a monaural, binaural-monotic, or dichotic nature. Recording can be controlled manually or by using the cue-interrupt circuit in the synthesizer interface. For the latter, an oscillator circuit is used to record tones on one or both channels of the tape at precise temporal intervals. As the cue-interrupt circuit senses each tone, synthesis is initiated and the tone is erased.

The amplifier/speaker unit is used to monitor the output of the synthesizer; it also allows monitoring during recording or playback of magnetic tape.

Software:

The primary software system used with the computer is the Disk Monitor System, written by DEC for its PDP-8 series of small computers (2). This is a file-oriented library system that allows the user to save programs and data on the DF32 disc storage unit and recall the information for use at other times. The system operates using commands which the user types on the teletype keyboard. Programs or data are given a name and are stored as files on the disc by issuing a "save" command; these same files can then be recalled by issuing a "call" command. Since these operations are performed at high speed, programs or data currently in the computer memory can be changed very rapidly.

Two programming languages used for controlling the speech synthesizer have been developed from the FOCAL language written by Richard Merrill of DEC (3). FOCPSY, developed by Steven Mullen, included a number of commands which allowed the user to control the speech synthesizer directly in a simple manner. Since these commands were an integral part of the language, they could be incorporated into any program written in the FOCPSY language, providing a powerful and flexible tool for speech synthesis (Mullen, 1970). The present language, FOCALAY, developed by Ronald Howes, uses a number of overlays which allow the user to modify the language for a specific purpose. Each overlay[1] has a unique set of commands designed for one experimental purpose. The overlay used in speech synthesis is called FP71, and has been used in several programs which allow speech synthesis on the system.

Programs:

VOXF. Mullen first developed this program for the input, output, and editing of control parameters. For each time frame the program requests values for each parameter (which the user supplies at the teletype) and stores them in the proper portion of the buffer. Corrections or changes may be made on values in any specified time frame; however, values of parameters of the time frame must be typed again. The experimenter may listen to the utterance at any time by using the "TALK" command. He may also control the nature of the synthesis at any time by specifying values for the duration of time frames (in milliseconds), the number of repetitions of the utterance, the delay between repetitions (in seconds), and the portion of the utterance in the buffer to be synthesized (specified by the first and last time frames). The experimenter may store any portion of the utterance on punched paper tape, and may read this back into the buffer for synthesis or editing at a later time. A "CUE" command is used for recording utterances at precise time intervals on magnetic tape and for synchronization of utterances on the two channels for dichotic recordings. (This command controls the cue interrupt circuit of the synthesizer interface described above.)

A recent change in the program allows the duration of each time frame in the utterance to be varied independently. This allows the "grain" of the synthesis to correspond to the rate of change of acoustic features.

VOX. Howes has written an expanded version of VOXF in machine language. Since it bypasses the FOCAL package, computation is faster, programming is more flexible, and more space is available

---

[1] An overlay is a short modification to a basic program which may be used to change the program by loading the overlay directly "over" the original program as it stands in the computer's memory.

in core for storage of synthesis control parameters. Editing is facilitated: Single values in specific time frames may be directly changed, a repeated value for a parameter in a succession of time frames can be stored by a single request, time frames may be inserted or deleted at any point, and a set of parameter values may be repeated in a series of time frames by a single request. VOX also allows a teletype listing of the values in the parameter buffer, an invaluable aid to editing. A parity-checking routine has been incorporated into the input/output functions; this allows most errors in the punching and reading of paper tapes to be detected. However, since VOX is written in machine language rather than FOCALAY, people who use the program find it more difficult to understand and revise.

VOXY. This FOCALAY program continuously outputs a single set of parameter values to the synthesizer. Any value may be altered directly from the teletype. This allows the experimenter to listen to particular resonances singly or in combination. The program is useful when synthesizing continuous speech sounds (such as vowels and fricatives) and when a quick check on the synthesizer's performance is desired.

SIN. Robert Verbrugge has developed a program to synthesize General American English by rule. The program is patterned after a synthesis-by-rule program designed by Holmes, Mattingly, and Shearme (1964). A phoneme table (stored in 2040 words on the disk) carries specifications of stressed and unstressed phoneme duration, type of excitation (pulse or noise source), steady-state values for each resonance frequency and amplitude, and parameters used in calculating transitions at phoneme boundaries. Stops, fricatives, affricates, and diphthongs are represented by two or more phoneme-parts in the table; some allophonic variants have separate representations. The input to the program is the string of phonemes for the desired utterance, each specified for stress level and pitch; durational values different from those in the table may also be supplied. The synthesis program (occupying about 3400 words of memory) accesses the necessary parameter values from the phoneme table, calculates the synthesizer control values, stores them in the parameter buffer, and outputs them to the synthesizer on command. The buffer can store parameter values for about three seconds of speech, with a 10 msec. time frame.

Several words, phrases, sentences, and bars of singing have been synthesized. The quality of speech from this first program (Original SIN) is relatively poor. The program is presently being revised to utilize additional components of the synthesizer as follows: (1) The nasal resonance circuit in nasal synthesis (now simulated with the lowest vocal resonance), (2) the wide-band fricative circuits in fricative synthesis (now simulated with noise excitation through the vocal resonances), (3) mixed source excitation in synthesizing voiced fricatives (now simulated by juxtaposing pulsed and noise segments). In addition, rules

for calculating allophonic variations and pitch contours, similar to the rules used by Mattingly (1968) are being incorporated. These revisions should substantially improve the quality of utterances synthesized by rule and provide a far richer context for studies of the perceptual cues for speech.

II. Infant Testing Laboratory

The facility for testing infants' auditory perception at Minnesota is modeled after Peter Eimas's laboratory at Brown University. Presentation of auditory stimuli is contingent on non-nutritive sucking by the infant. The infant is placed on a semi-reclining infant seat which is attached to a table. An AR 2AX speaker over which stimuli are presented is placed in front of the infant and a video camera for remote monitoring is focused on the infant. An autoclaved pacifier nipple connected by a plastic hose to a Stratham P23 "V" pressure transducer is held in the infant's mouth by a research assistant. The transducer output feeds into a two-channel Beckman R.S. dynagraph, which provides a graphic record of all sucking behavior. The output of the dynagraph is also fed directly into a PDP-12 computer system for analysis. Criterion amplitude sucks operate an electronic timer and switch which control the output of an Ampex 602-2 tape recorder driving the speaker.

From an adjacent room, the experimenter operates switches that control the criterion level of sucking responses, the onset of the stimulus control system, and the choice of the stimuli presented to the infant.

This facility is also used in studies on infant perception which use the heart rate measure described in the research section. A continuous record of heart response is obtained with the dynagraph via contact electrodes placed on the subject's body. The dynagraph output is fed into the PDP-12 computer through the six analogue input channels or is stored on tape with a Vetter FM recording adapter.

This facility was developed jointly with funds from the OE grant and from the Program Project Grant of the Institute for Child Development. Its use is shared with members of the Institute.

REFERENCES

1. PDP-8/L Users' Handbook. Maynard, Massachusetts: Digital Equipment Corporation, 1968.

2. Disk Monitor System. DEC-08-SDAB-D, 1969.

3. Focal-8. DEC-08-AJAD-D, 1969.

Glace, D.A. A parallel resonance synthesizer for speech research. Paper read at 75th meeting of Acoustical Society of America, May, 1968.

Holmes, J.N., Mattingly, I. G., & Shearme, J. N. Speech synthesis by rule. Language and Speech, 1964, 7, 127-143.

Mattingly, I. G. Synthesis by rule of General American English. Doctoral dissertation, Yale University. In Supplement to Status Report on Speech Research, Haskins Laboratories, New Haven, Connecticut, 1968.

Mullen, S. L. A constructive analysis of PDP-8 programming systems as applied to computer-controlled experimentation . Unpublished manuscript, Center for Research in Human Learning, July, 1970.

APPENDIX I

DISSEMINATION ACTIVITIES

Oral Presentations

1. Dr. James J. Jenkins. "The Psychology of Speech Perception." Lecture-demonstration given as part of the series: Varieties of Academic Experience, Department of Chemical Engineering, University of Minnesota, February, 1970.

2. Dr. James Jenkins & Winifred Strange. "The Psychology of Speech Perception." Lecture-demonstration to members of the Department of Electrical Engineering, University of Minnesota, April, 1970.

3. Winifred Strange. Lectures in academic courses: Psychology of Language, May, 1970, and Acoustic Phonetics, May, 1970.

4. Winifred Strange. "The perception of voicing: Cross language studies." Lecture to members of the Department of Linguistics, February, 1971.

5. Dr. James Jenkins. "Language and Speech." A 40-minute film prepared for Introductory Psychology, July, 1971.

6. Dr. James Jenkins. "Some possible parallels between speech perception and verbal learning." Lecture delivered at the Eastern Verbal Investigators League, Smith College, Amherst, Massachusetts, October, 1971.

7. Dennis Doty. "Speech perception in children and infants." Lecture delivered at:

   Center for Research in Human Learning, University of Minnesota, February, 1972
   New York University, New York, February, 1972
   Carleton University, Ottawa, Canada, February, 1972
   University of Illinois, Chicago Circle, February, 1972

Publications

Barclay, J.R. Non-categorical perception of a voiced stop: A replication. Perception and Psychophysics, 1972, 11-4, 269-273.

Doty, D. Training ten- and eleven-year-olds to discriminate within phoneme boundaries along the voicing continuum. Perception and Psychophysics, in press, 1972.

Strange, W., & Halwes, T. Confidence ratings in speech perception research: Evaluation of an efficient technique for discrimination testing. Perception and Psychophysics, 1971, 9, 182-186.

APPENDIX II

PERSONNEL

Dr. James J. Jenkins (Principal Investigator): Professor of Psychology, University of Minnesota; Director of Research, Center for Research in Human Learning (CRHL).

J. Richard Barclay: NIH Predoctoral Fellow, 1968-1971.

Dennis Doty: CRHL Predoctoral Fellow, 1968-1972.

Winifred Strange: NIH Predoctoral Fellow, 1969-1971
CRHL Research Fellow, 1971 to present.

Robert Verbrugge: Danforth Fellow, 1969 to present.

Patricia Yonas: Instructor, Institute of Child Development.

Ronald Howes: Undergraduate Research Assistant, 1971 to present.

Steven Mullen: Graduate Research Assistant, 1969-1970.